Web-based Supplementary Materials for for "Penalized Generalized Estimating Equations for High-dimensional Longitudinal Data Analysis"

by Lan Wang, Jianhui Zhou and Annie Qu

Web Appendix A: A Remark

It is easy to see that equation (2) in Section 2.1 follows directly from equation (1) when the marginal distribution of Y_{ij} is from a canonical exponential family(common assumption for GEE): $\frac{\partial \boldsymbol{\mu}_i(\boldsymbol{\beta}_n)}{\partial \boldsymbol{\beta}_n} = \phi \mathbf{A}_i \mathbf{X}_i$ and $\mathbf{V}_i = \mathbf{A}_i^{1/2}\mathbf{R}\mathbf{A}_i^{1/2}$.

Web Appendix B: Proof of Theorem 1

Throughout the proof, we use C to denote a generic constant, which is independent of n and may vary from line to line.

Let

$$\overline{\mathbf{S}}_n(\boldsymbol{\beta}_n) = n^{-1}\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-1/2}(\boldsymbol{\beta}_n)(\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta}_n)).$$

We write $\overline{\mathbf{S}}_n(\boldsymbol{\beta}_n) = (\overline{S}_{n1}(\boldsymbol{\beta}_n), \ldots, \overline{S}_{np_n}(\boldsymbol{\beta}_n))^T$, where $\overline{S}_{nk}(\boldsymbol{\beta}_n) = \mathbf{e}_k^T\overline{\mathbf{S}}_n(\boldsymbol{\beta}_n)$, and $\mathbf{e}_k$ is a p_n dimensional basis vector with the kth element being one and all the other elements being zero, $1 \le k \le p_n$. In the following, we first present a useful lemmas.

Lemma 0.1 *Assume conditions (A1)-(A7) in Section 4 hold, then*

$$\frac{\partial \overline{S}_{nk}(\boldsymbol{\beta}_n)}{\partial \boldsymbol{\beta}_n^T} = \mathbf{H}_{nk}(\boldsymbol{\beta}_n) + \mathbf{E}_{nk}(\boldsymbol{\beta}_n) + \mathbf{G}_{nk}(\boldsymbol{\beta}_n), \tag{1}$$

where

$$\mathbf{H}_{nk}(\boldsymbol{\beta}_n) = -n^{-1}\sum_{i=1}^{n}\mathbf{e}_k^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\mathbf{X}_i,$$

$$
\begin{aligned}
\mathbf{E}_{nk}(\boldsymbol{\beta}_n) &= -(2n)^{-1}\sum_{i=1}^{n}\mathbf{e}_k^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-3/2}(\boldsymbol{\beta}_n)\mathbf{C}_i(\boldsymbol{\beta}_n)\mathbf{F}_i(\boldsymbol{\beta}_n)\mathbf{X}_i,\\
\mathbf{G}_{nk}(\boldsymbol{\beta}_n) &= (2n)^{-1}\sum_{i=1}^{n}\mathbf{e}_k^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_n)\mathbf{F}_i(\boldsymbol{\beta}_n)\mathbf{J}_i(\boldsymbol{\beta}_n)\mathbf{X}_i,
\end{aligned}
$$

with

$$
\begin{aligned}
\mathbf{C}_i(\boldsymbol{\beta}_n) &= diag(Y_{i1}-\mu_{i1}(\boldsymbol{\beta}_n),\ldots,Y_{im}-\mu_{im}(\boldsymbol{\beta}_n)),\\
\mathbf{F}_i(\boldsymbol{\beta}_n) &= diag(\ddot{\mu}(\mathbf{X}_{i1}^T\boldsymbol{\beta}_n),\ldots,\ddot{\mu}(\mathbf{X}_{im}^T\boldsymbol{\beta}_n)),\\
\mathbf{J}_i(\boldsymbol{\beta}_n) &= diag(\overline{\mathbf{R}}^{-1}\mathbf{A}_i^{-1/2}(\boldsymbol{\beta}_n)(\mathbf{Y}_i-\boldsymbol{\mu}_i(\boldsymbol{\beta}_n))).
\end{aligned}
$$

In the above, for $\mathbf{a}=(a_1,\ldots,a_m)^T$, $diag(a_1,\ldots,a_m)$ *and* $diag(\mathbf{a})$ *both denote an* $m\times m$ *diagonal matrix with diagonal entries* $a_1,\ldots,a_m$.

The derivation of Lemma 1 can be found in Wang (2011). Lemma 2 below can be found in van der Vaart and Wellner (1996, Lemma 2.2.11).

Lemma 0.2 *(Bernstein's inequality) Let* $Z_1,\ldots,Z_n$ *be independent random variables with mean zero and satisfy*

$$
E|Z_i|^l \le l!M^{l-2}V_i/2
$$

for every $l\ge 2$ *and all* i *and some positive constants* M *and* V_i. *Then*

$$
P(|Z_1+\ldots+Z_n|>t)\le 2\exp\left(-\frac{1}{2}\frac{t^2}{V+Mt}\right),
$$

for $V>V_1+\ldots+V_n$.

Proof of Theorem 1. We prove the theorem by construction. Let $\widehat{\boldsymbol{\beta}}_n=(\widehat{\boldsymbol{\beta}}_{n1}^T,\mathbf{0}^T)^T$ be the oracle estimator. We'll show that $\widehat{\boldsymbol{\beta}}_n$ satisfies properties (1)-(3). Properties (2) and (3) follow from the definition of $\widehat{\boldsymbol{\beta}}_n$ and the results in Wang (2011). In what follows, we verify that $\widehat{\boldsymbol{\beta}}_n$ satisfies (6) and (7).

Proof of (6). We have $S_{nj}(\widehat{\boldsymbol{\beta}}_n) = 0$, $j = 1, \ldots, s_n$, from the definition of $\widehat{\boldsymbol{\beta}}_n$. It thus suffices to show that $P(|\widehat{\beta}_{nj}| \geq a\lambda_n,\ j = 1, \ldots, s_n) \to 1$, as this implies the penalty function to be zero with probability approaching one. Note that $\min_{1 \leq j \leq s_n} |\widehat{\beta}_{nj}| \geq \min_{1 \leq j \leq s_n} |\beta_{n0j}| - \max_{1 \leq j \leq s_n} |\beta_{n0j} - \widehat{\beta}_{nj}| \geq \min_{1 \leq j \leq s_n} |\beta_{n0j}| - ||\boldsymbol{\beta}_{n10} - \widehat{\boldsymbol{\beta}}_{n10}||$. From Wang (2010),

$$||\boldsymbol{\beta}_{n10} - \widehat{\boldsymbol{\beta}}_{n10}|| = \sqrt{s_n/n}. \tag{2}$$

Therefore, we have

$$\begin{aligned} & P(\min_{1 \leq j \leq s_n} |\beta_{n0j}| - ||\boldsymbol{\beta}_{n10} - \widehat{\boldsymbol{\beta}}_{n10}|| \geq a\lambda_n) \\ = \ & P(||\boldsymbol{\beta}_{n10} - \widehat{\boldsymbol{\beta}}_{n10}|| \leq \min_{1 \leq j \leq s_n} |\beta_{n0j}| - a\lambda_n) \to 1 \end{aligned}$$

since $\min_{1 \leq j \leq s_n} |\beta_{n0j}|/\lambda \to \infty$ and $||\boldsymbol{\beta}_{n10} - \widehat{\boldsymbol{\beta}}_{n10}|| = o(\lambda_n)$. Thus $P(\min_{1 \leq j \leq s_n} |\widehat{\boldsymbol{\beta}}_{nj}| \geq a\lambda_n) \to 1$ as $n \to \infty$.

Proof of (7). We have $\widehat{\beta}_{nk} = 0$, thus $q_{\lambda_n}(\widehat{\beta}_{nk})\mathrm{sign}(\widehat{\beta}_{nk}) = 0$, for $k = s_n + 1, \ldots, p_n$, from the definition of $\widehat{\boldsymbol{\beta}}$. To prove (7), it suffices to verify that

$$P\Big(\max_{s_n+1 \leq k \leq p_n} |S_{nk}(\widehat{\boldsymbol{\beta}}_n)| \leq \frac{\lambda_n}{\log(n)}\Big) \to 1. \tag{3}$$

The statement in (3) is implied by

$$P\Big(\max_{s_n+1 \leq k \leq p_n} |S_{nk}(\widehat{\boldsymbol{\beta}}_n) - \overline{S}_{nk}(\widehat{\boldsymbol{\beta}}_n)| > \frac{\lambda_n}{2\log(n)}\Big) \to 0. \tag{4}$$

$$P\Big(\max_{s_n+1 \leq k \leq p_n} |\overline{S}_{nk}(\widehat{\boldsymbol{\beta}}_n)| > \frac{\lambda_n}{2\log(n)}\Big) \to 0. \tag{5}$$

The left side of (4) is bounded from above by

$$
\begin{aligned}
& P\Big(\max_{s_n+1\le k\le p_n} n^{-1}\sum_{i=1}^{n}\Big|\mathbf{e}_k^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\widehat{\boldsymbol\beta}_n)[\widehat{\mathbf{R}}^{-1}-\overline{\mathbf{R}}^{-1}]\mathbf{A}_i^{-1/2}(\widehat{\boldsymbol\beta}_n)(\mathbf{Y}_i-\boldsymbol\mu_i(\widehat{\boldsymbol\beta}_n))\Big| > \frac{\lambda_n}{2\log(n)}\Big) \\
\le\ & P\Big(\max_{s_n+1\le k\le p_n} n^{-1}\sum_{i=1}^{n}||\mathbf{e}_k^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\widehat{\boldsymbol\beta}_n)||\cdot||\widehat{\mathbf{R}}^{-1}-\overline{\mathbf{R}}^{-1}||\cdot||\mathbf{A}_i^{-1/2}(\widehat{\boldsymbol\beta}_n)(\mathbf{Y}_i-\boldsymbol\mu_i(\widehat{\boldsymbol\beta}_n))|| \\
& \qquad > \frac{\lambda_n}{2\log(n)}\Big) \\
\le\ & P\Big(||\widehat{\mathbf{R}}^{-1}-\overline{\mathbf{R}}^{-1}||n^{-1}\sum_{i=1}^{n}\Big(\max_{s_n+1\le k\le p_n}||\mathbf{e}_k^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\widehat{\boldsymbol\beta}_n)||\Big)||\mathbf{A}_i^{-1/2}(\widehat{\boldsymbol\beta}_n)(\mathbf{Y}_i-\boldsymbol\mu_i(\widehat{\boldsymbol\beta}_n))|| \\
& \qquad > \frac{\lambda_n}{2\log(n)}\Big) \\
\le\ & P\Big(n^{-1}\sum_{i=1}^{n}||\epsilon_i(\widehat{\boldsymbol\beta}_n)|| > \frac{\lambda_n\sqrt{n}}{2\sqrt{s_n}\log(n)}\Big) \\
\le\ & C\frac{n^{-1}\sum_{i=1}^{n}E(||\epsilon_i(\widehat{\boldsymbol\beta}_n)||)\sqrt{s_n}\log n}{\lambda_n\sqrt{n}} = O\big(\frac{\sqrt{s_n}\log n}{\lambda_n\sqrt{n}}\big) = o(1),
\end{aligned}
$$

where the third inequality follows from conditions (A1), (A4) and (A6), and it follows from condition (A7) that $\sqrt{s_n}\log n/(\sqrt{n}\lambda_n)\to 0$. To prove (5), we consider the following Taylor expansion:

$$
\overline{S}_{nk}(\widehat{\boldsymbol\beta}_n) = \overline{S}_{nk}(\boldsymbol\beta_{n0}) + \frac{\partial\overline{S}_{nk}(\boldsymbol\beta_n)}{\partial\boldsymbol\beta_n^T}(\widehat{\boldsymbol\beta}_n-\boldsymbol\beta_{n0}) + (\widehat{\boldsymbol\beta}_n-\boldsymbol\beta_{n0})^T\frac{\partial^2\overline{S}_{nk}(\boldsymbol\beta_n^*)}{\partial\boldsymbol\beta_n\partial\boldsymbol\beta_n^T}(\widehat{\boldsymbol\beta}_n-\boldsymbol\beta_{n0}), \quad (6)
$$

where $\boldsymbol\beta_n^*$ is between $\boldsymbol\beta_{n0}$ and $\widehat{\boldsymbol\beta}_n$. Let $\nabla_k(\boldsymbol\beta_n)$ denote $\frac{\partial\overline{S}_{nk}(\boldsymbol\beta_n)}{\partial\boldsymbol\beta_n^T}$ and let $\mathbf{D}_k(\boldsymbol\beta_n)$ denote $\frac{\partial^2\overline{S}_{nk}(\boldsymbol\beta_n)}{\partial\boldsymbol\beta_n\partial\boldsymbol\beta_n^T}$. Let $\nabla_{k1}(\boldsymbol\beta_n)$ be the subvector that consists of the first s_n elements of $\nabla_k(\boldsymbol\beta_n)$, and let $\mathbf{D}_{k1}(\boldsymbol\beta_n)$ denote the $s_n\times s_n$ submatrix in the upper-left corner of $\mathbf{D}_k(\boldsymbol\beta_n)$. Since $\widehat{\boldsymbol\beta}_n-\boldsymbol\beta_{n0} = ((\widehat{\boldsymbol\beta}_{n1}-\boldsymbol\beta_{n10})^T,\mathbf{0}^T)^T$, (6) becomes

$$
\overline{S}_{nk}(\widehat{\boldsymbol\beta}_n) = \overline{S}_{nk}(\boldsymbol\beta_{n0}) + \nabla_{k1}(\boldsymbol\beta_{n0})(\widehat{\boldsymbol\beta}_{n1}-\boldsymbol\beta_{n10}) + (\widehat{\boldsymbol\beta}_{n1}-\boldsymbol\beta_{n10})^T\mathbf{D}_{k1}(\boldsymbol\beta_n^*)(\widehat{\boldsymbol\beta}_{n1}-\boldsymbol\beta_{n10}).
$$

Note that

$$
P\Big(\max_{s_n+1\le k\le p_n}|\overline{S}_{nk}(\widehat{\boldsymbol\beta}_n)| > \frac{\lambda_n}{2\log(n)}\Big)
$$

$$\leq P\Big(\max_{s_n+1\leq k\leq p_n}|\overline{S}_{nk}(\boldsymbol{\beta}_{n0})| > \frac{\lambda_n}{6\log(n)}\Big) + P\Big(\max_{s_n+1\leq k\leq p_n}|\nabla_{k1}(\boldsymbol{\beta}_{n0})(\widehat{\boldsymbol{\beta}}_{n1}-\boldsymbol{\beta}_{n10})| > \frac{\lambda_n}{6\log(n)}\Big)$$
$$+P\Big(\max_{s_n+1\leq k\leq p_n}|(\widehat{\boldsymbol{\beta}}_{n1}-\boldsymbol{\beta}_{n10})^T\mathbf{D}_{k1}(\boldsymbol{\beta}_n^*)(\widehat{\boldsymbol{\beta}}_{n1}-\boldsymbol{\beta}_{n10})| > \frac{\lambda_n}{6\log(n)}\Big) = I_{n1}+I_{n2}+I_{n3}.$$

Thus (5) is implied by $I_{ni}=o(1)$, $i=1,2,3$, which is verified below.

First, we'll show that $I_{n1}=o(1)$. We have

$$I_{n1} \leq \sum_{k=s_n+1}^{p_n} P\Big(|\overline{S}_{nk}(\boldsymbol{\beta}_{n0})| > \frac{\lambda_n}{6\log(n)}\Big).$$

We can write $\overline{S}_{nk}(\boldsymbol{\beta}_{n0}) = n^{-1}\sum_{i=1}^n Z_i$, where $Z_i = \mathbf{e}_k^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})\overline{\mathbf{R}}^{-1}\boldsymbol{\epsilon}_i(\boldsymbol{\beta}_{n0})$ are independent mean zero random variables. Note that $\forall\ l\geq 2$, we have

$$\begin{aligned}
E|Z_i|^l &\leq E[||\mathbf{e}_k^T\mathbf{X}_i^T\mathbf{A}_i^{1/2}(\boldsymbol{\beta}_{n0})\overline{\mathbf{R}}^{-1}||^l\cdot||\boldsymbol{\epsilon}_i(\boldsymbol{\beta}_{n0})||^l]\\
&\leq C^lE\Big[||\boldsymbol{\epsilon}_i||^l\Big] = C^lE\Big[\Big(\sum_{j=1}^m \epsilon_{ij}^2\Big)^{l/2}\Big]\\
&\leq C^lm^{l/2-1}\sum_{j=1}^m E|\epsilon_{ij}|^l \leq C^lm^{l/2-1}\sum_{j=1}^m l!M_2^{-l}\mathrm{E}(\exp(M_2|\epsilon_{ij}|))\\
&\leq C^lm^{l/2-1}ml!M_3 \leq l!M^{l-2}\delta/2,
\end{aligned}$$

for some constants $M>0$ and $\delta>0$. In the above derivation, the second inequality follows from conditions (A1), (A4) and (A6), the third inequality follows from the fact $|\sum_{i=1}^m a_i|^p \leq m^{p-1}\sum_{i=1}^m|a_i|^p$ for $m\geq 1$ and $p\geq 1$ (a result of Jensen's inequality); the fourth inequality follows from the Taylor expansion of the exponential function and the second last inequality follows from condition (A5). Thus the Z_i satisfy the conditions of Bernstein's inequality. By Lemma 2, we have

$$\begin{aligned}
P\Big(|\overline{S}_{nk}(\boldsymbol{\beta}_{n0})| > \frac{\lambda_n}{6\log(n)}\Big) &\leq 2\exp\Big[-\frac{1}{2}\frac{n^2\lambda_n^2/(36(\log n)^2)}{n\delta+M^*n\lambda_n/(6\log n)}\Big]\\
&\leq 2\exp\Big[-C\frac{n\lambda_n^2}{(\log n)^2}\Big].
\end{aligned}$$

Therefore

$$I_{n1} \leq 2\exp\Big[\log p_n - C\frac{n\lambda_n^2}{(\log n)^2}\Big] = o(1),$$

because $\log p_n = o(n\lambda_n^2/(\log n)^2)$ and $n\lambda_n^2/(\log n)^2 \to \infty$ as $n \to \infty$ by condition (A7). This verifies $I_{n1} = o(1)$.

Next we'll prove that $I_{n2} = o(1)$. We have

$$\begin{aligned}
I_{n2} &= P\Big(\max_{s_n+1\leq k\leq p_n} |\nabla_{k1}(\boldsymbol{\beta}_{n0})(\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n10})| > \frac{\lambda_n}{6\log(n)}\Big) \\
&= P\Big(\max_{s_n+1\leq k\leq p_n} |\nabla_{k1}(\boldsymbol{\beta}_{n0})(\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n10})| > \frac{\lambda_n}{6\log(n)},\ ||\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n10}|| \leq \sqrt{s_n/n}\log n\Big) \\
&\quad + P(||\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n10}|| > \sqrt{s_n/n}\log n) \\
&\leq P(\max_{s_n+1\leq k\leq p_n} ||\nabla_{k1}(\boldsymbol{\beta}_{n0})|| > \frac{\lambda_n\sqrt{n}}{6\sqrt{s_n}(\log n)^2}) + o(1) \\
&\leq P(\max_{s_n+1\leq k\leq p_n} ||\mathbf{H}_{nk1}(\boldsymbol{\beta}_{n0})|| > \frac{\lambda_n\sqrt{n}}{18\sqrt{s_n}(\log n)^2}) \\
&\quad + P(\max_{s_n+1\leq k\leq p_n} ||\mathbf{E}_{nk1}(\boldsymbol{\beta}_{n0})|| > \frac{\lambda_n\sqrt{n}}{18\sqrt{s_n}(\log n)^2}) \\
&\quad P(\max_{s_n+1\leq k\leq p_n} ||\mathbf{G}_{nk1}(\boldsymbol{\beta}_{n0})|| > \frac{\lambda_n\sqrt{n}}{18\sqrt{s_n}(\log n)^2}) + o(1) \\
&= I_{n21} + I_{n22} + I_{n23} + o(1),
\end{aligned}$$

where $\mathbf{H}_{nk1} = (H_{nk1}, \ldots, H_{nks_n})^T$ denotes the subvector of $\mathbf{H}_{nk}$ which consists its first s_n elements, $\mathbf{E}_{nk1}$ and $\mathbf{G}_{nk1}$ are defined similarly, the first inequality uses (2), the second inequality uses Lemma 1, and the definition of I_{n2i} $(i = 1, 2, 3)$ should be clear from the context. To evaluate I_{n21}, we observe that

$$\begin{aligned}
I_{n21} &\leq P(\max_{s_n+1\leq k\leq p_n} ||\mathbf{H}_{nk1}(\boldsymbol{\beta}_{n0})||^2 > C\frac{n\lambda_n^2}{s_n(\log n)^4}) \\
&\leq P\Big(\max_{s_n+1\leq k\leq p_n} |||\mathbf{H}_{nk1}(\boldsymbol{\beta}_{n0})||^2 - E||\mathbf{H}_{nk1}(\boldsymbol{\beta}_{n0})||^2| \\
&\qquad + \max_{s_n+1\leq k\leq p_n} E||\mathbf{H}_{nk1}(\boldsymbol{\beta}_{n0})||^2 > C\frac{n\lambda_n^2}{s_n(\log n)^4}\Big).
\end{aligned}$$

By conditions (A1), (A4) and (A6), $|H_{nkj}(\boldsymbol{\beta}_{n0})|$ is uniformly bounded by a positive constant. Thus $\max_{s_n+1\le k\le p_n} E||\mathbf{H}_{nk1}(\boldsymbol{\beta}_{n0})||^2 = \max_{s_n+1\le k\le p_n} E(\sum_{j=1}^{s_n} H_{nkj}^2(\boldsymbol{\beta}_{n0})) \le Cs_n$. Since $s_n^2(\log n)^4 = o(n\lambda_n^2)$ and $p_n s_n^3(\log n)^8/(n^2\lambda_n^4) = o(1)$by condition (A7), for n sufficiently large, we have

$$
\begin{aligned}
I_{n21} &\le P\Big(\max_{s_n+1\le k\le p_n} |||\mathbf{H}_{nk1}(\boldsymbol{\beta}_{n0})||^2 - E||\mathbf{H}_{nk1}(\boldsymbol{\beta}_{n0})||^2| > \frac{C}{2}\frac{n\lambda_n^2}{s_n(\log n)^4}\Big) \\
&\le \sum_{k=s_n+1}^{p_n} P\Big(|||\mathbf{H}_{nk1}(\boldsymbol{\beta}_{n0})||^2 - E||\mathbf{H}_{nk1}(\boldsymbol{\beta}_{n0})||^2| > \frac{C}{2}\frac{n\lambda_n^2}{s_n(\log n)^4}\Big) \\
&\le C\sum_{k=s_n+1}^{p_n} \frac{E[\sum_{j=1}^{s_n}(H_{nkj}^2(\boldsymbol{\beta}_{n0}) - E(H_{nkj}^2(\boldsymbol{\beta}_{n0}))]^2 s_n^2(\log n)^8}{n^2\lambda_n^4} \\
&= O(p_n s_n^3(\log n)^8/(n^2\lambda_n^4)) = o(1),
\end{aligned}
$$

where the third inequality applies Markov's inequality. Similarly as above, we can show that $I_{n22} = o(1)$ and $I_{n23} = o(1)$. And this verifies $I_{n2} = o(1)$.

Finally, we verify that $I_{n3} = o(1)$. We have

$$
\begin{aligned}
I_{n3} &\le P\Big(\max_{s_n+1\le k\le p_n} |(\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n10})^T\mathbf{D}_{k1}(\boldsymbol{\beta}_n^*)(\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n10})| > \frac{\lambda_n}{6\log(n)}\Big) \\
&\le P\Big(\max_{s_n+1\le k\le p_n} |(\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n10})^T\mathbf{D}_{k1}(\boldsymbol{\beta}_n^*)(\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n10})| > \frac{\lambda_n}{6\log(n)}, \\
&\qquad ||\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n10}|| \le \sqrt{s_n/n}\log n\Big) + P(||\widehat{\boldsymbol{\beta}}_{n1} - \boldsymbol{\beta}_{n10}|| > \sqrt{s_n/n}\log n) \\
&\le \sum_{k=s_n+1}^{p_n} P\Big(\mathrm{tr}(\mathbf{D}_{k1}(\boldsymbol{\beta}_n^*)) > \frac{n\lambda_n}{s_n(\log n)^3}\Big) + o(1) \\
&\le C\sum_{k=s_n+1}^{p_n} \frac{E[\mathrm{tr}(\mathbf{D}_{k1}(\boldsymbol{\beta}_n^*)^2]s_n^2(\log n)^6}{n^2\lambda_n^2} + o(1),
\end{aligned}
$$

where the third inequality uses (2) and the last inequality applies Markov's inequality.. Note that

$$
E[\mathrm{tr}(\mathbf{D}_{k1}(\boldsymbol{\beta}_n^*)^2] = E\Big[\sum_{j=1}^{s_n} \frac{\partial^2 \overline{S}_{nk}}{\partial\beta_{nj}^2}(\boldsymbol{\beta}_n^*)\Big]^2 \le Cs_n^2,
$$

uniformly in k, by conditions (A1), (A4), (A5) and (A6). Thus $I_{n3} = O(p_n s_n^4 (\log n)^6/(n^2\lambda_n^2)) + o(1) = o(1)$ since $p_n s_n^4 (\log n)^6/(n^2\lambda_n^2) = o(1)$ by condition (A7).

Putting the above together, we have proved Theorem 1. □

Additional References

van der Vaart, A. and Wellner, J. (1996) Weak convergence and empirical processes: with applications to statistics. Springer: New York.